DOCUMENT RESUME

ED 356 481 CS 213 772

| | |
|---|---|
| AUTHOR | Peters, Joanne K. A. |
| TITLE | Certain Uncertainties: New Literacy and the Evaluation of Student Writing. NCTE Concept Papers Series. Concept Paper No. 5. |
| INSTITUTION | National Council of Teachers of English, Urbana, Ill. |
| PUB DATE | 92 |
| NOTE | 57p. |
| AVAILABLE FROM | National Council of Teachers of English, 1111 W. Kenyon Road, Urbana, IL 61801 (Stock No. 05262). |
| PUB TYPE | Viewpoints (Opinion/Position Papers, Essays, etc.) (120) |
| EDRS PRICE | MF01/PC03 Plus Postage. |
| DESCRIPTORS | Elementary Secondary Education; Higher Education; *Politics of Education; *Student Evaluation; *Teacher Attitudes; *Writing Evaluation |
| IDENTIFIERS | Alternative Assessment; Educational Issues; *English Teachers; Literacy as a Social Process; *Student Centered Assessment; Text Factors |

ABSTRACT

This paper explores how teachers who choose the New Literacy approach to the teaching of reading and writing reconcile the conflict between their need for personal pedagogical integrity and public demands for accountability. The paper's first section describes one teacher's changing conceptions of evaluation of student writing. The second section explores the long-standing belief that it is possible to assess written work according to some absolute, reliable, and valid standard, and argues that this belief is a myth. The next section argues that evaluation of student work is a function of a teacher's perception of his or her professional role and purpose, which is in turn linked to beliefs in how meaning is made from and with text. The problem for the teacher is the choice between a supportive role with textual meaning as open to negotiation versus an authoritative role with textual meaning largely fixed. The fourth section of the paper explores ways in which evaluation can support New Literacy models of the reading and writing process while appeasing public demands for quantified measures of student writing ability. The final section reflects on the politics of assessment and the challenges teachers face from a public demanding accountability in the form of reductive grades. (Contains 52 references.) (SR)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED356481

# NCTE Concept Papers

Concept Paper No. 5

## Certain Uncertainties: New Literacy and the Evaluation of Student Writing

Joanne K. A. Peters
Sisler High School, Winnipeg, Manitoba

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Maxwell

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

NCTE

National Council of Teachers of English
1111 Kenyon Road, Urbana, Illinois 61801

CS213773

BEST COPY AVAILABLE

**NCTE**
**Concept Paper Series**

Concept Paper No. 5

# Certain Uncertainties: New Literacy and the Evaluation of Student Writing

Joanne K. A. Peters
Sisler High School, Winnipeg, Manitoba

Copyright © 1992 by the National Council of Teachers of English. All rights reserved. Direct reprint requests to the Copyright Clearance Center, 27 Congress Street, Salem, Massachusetts 01970.

It is the policy of NCTE in its journals and other publications to provide a forum for the open discussion of ideas concerning the content and the teaching of English and the language arts. Publicity accorded to any particular point of view does not imply endorsement by the NCTE Executive Committee, the Board of Directors, or the membership at large, except in announcements of policy, where such endorsement is clearly specified.

NCTE Editorial Board: Keith Gilyard; Ronald Jobe; Joyce Kinkead; Louise W. Phelps; Gladys Veidemanis; Charles Suhor, chair, *ex officio;* Michael Spooner, *ex officio.*

NCTE Stock Number 05262

National Council of Teachers of English
1111 Kenyon Road, Urbana, Illinois 61801

**Certain Uncertainties:**

**New Literacy and the Evaluation of Student Writing**

## Remembrance of Things Past, Reflections on Things Present

Fifteen years ago, I graded a high school student's writing for the first time. I was a fourth-year Honors English student intending to become an English teacher. Knowing of my plans, and of a university student's constant need for cash, my former Grade 12 English teacher offered me part-time employment as a grader of high school English papers. Thus began my induction into the "common sense culture of English departments" (Ede, 1989, p. 149), where reading and grading volumes of student writing is both the badge of honor and cross of martyrdom borne by teachers.

Grading was much simpler for me then. I did not believe that only one interpretation of a literary work was possible, but I was certain that the author's meaning resided deep within the text. The student writing an argumentative essay had to extract that meaning and explicate it in a logically organized but thoroughly artistic written response to the assigned topic or prompt.

With the naiveté (or arrogance) of the very young, I forgot (or failed to realize) that I had far more extensive knowledge of English literature in general, of the literary work under examination, and of the discourse conventions of argumentative essays than did the students whose papers I read and evaluated. Thus, I measured each piece of writing against a model text, a Platonic Ideal Essay (Brannon & Knoblauch, 1982, p. 159) which I fashioned in my mind. Anyone who wrote as well as I did in Grade 12 probably

deserved an A, anyone who merely retold the narrative of the piece deserved a C, and anyone whose writing was fragmented and painful received an F.

Seven years passed. I started to suffer misgivings about the process of evaluation practiced by me and my colleagues, about the types of writing products we graded, and about the fairness, consistency, and validity of it all. Other changes were happening, too: a new curriculum in English Language Arts was being introduced and implemented in Manitoba. It stressed student knowledge and practice of the process of writing and proposed a new role for the teacher, that of supportive coach and facilitator; both changes moved "students and teachers out of traditional patterns of classroom behavior" (Manitoba Education, 1987, p. 4). Intrigued by this reconception of English curriculum, I was psychologically and ideologically ripe for conversion to a new concept of literacy.

I learned the process of writing and how to practice it with my students. I learned to trust my intuitive belief that students should read literature other than the canon of classroom classics, written in forms other than novel, play, or poem. I learned that student text could be crafted in forms other than essays, in voices more compelling than the objectified third person, and I delighted at what resonated in my mind as I read their journals, dialogues, and stories. I had become both a student and a teacher of "the New Literacy" (Willinsky, 1990).

But converts, despite their enthusiasm and zeal, bring to their new faith the legacy of a previous ideological heritage. Every time my students completed their writing assignments, every time I faced a set of blank mark recorder sheets, and every time I stared at a forbidding stack of semester-end exams, I asked myself questions: How can I

respond in a manner which is supportive, yet renders a fair judgment? How am I to evaluate fairly the composing processes which have led to these very diverse products? And how am I supposed to reduce it all to a mark which purportedly tells the administration, the student, his or her parents, and other colleagues how effectively this kid has learned, both as an individual and as a member of a group of students? It seemed an impossible problem, and discussions with other teachers told me that I was not alone in my difficulty.

Difficult or disheartening as the task can be, grading, evaluation, or assessment--the process of examining student work for signs of growth or mastery of skill and content--is unlikely to disappear from the world of education. It is firmly embedded in educational tradition and, I suspect, within the whole of Western culture. But new concepts of English curriculum demand a reexamination of a very old question: "How can the effectiveness of learning experiences be evaluated?" (Tyler, 1949). The question is now even more compelling. Teachers, "their professional authority eroded and their curricular choices further determined by school administrators quick to call for test scores" (Willinsky, 1990, p. xvi), face considerable political pressure while attempting to justify curricular change.

New Literacy is not simply curricular innovation, new methodology and new content, valuable though that can be. It is also a reconceptualization of the purpose and intention behind literacy education and a renegotiation of the classroom contract between student and teacher. As John Willinsky (1990) explains,

> The New Literacy consists of those strategies in the teaching of reading and writing which attempt to shift the control of literacy from the teacher to the student; literacy is promoted in such programs as a social process with language that can from the very beginning extend the students' range of meaning and connection. (p. 8)

Clearly, both students and teachers enact new roles in the New Literacy classroom. In their own and their students' eyes, teachers become coaches, editors, agents, and publishers (Willinsky, 1990, pp. 9-10). But the public still sees teachers as judges and evaluators, wielding the authority of the red pencil and the gradebook. Opposing expectations and dichotomous roles create profound strain and raise difficult questions. My question is this: How do teachers who choose the New Literacy approach to the teaching of reading and writing undertake a similar change in their evaluation of student work? How are teachers to reconcile the conflict between their need for personal pedagogical integrity and public demands for accountability?

I cannot pretend that I have a single, simple answer, but by exploring several key issues in the evaluation of written work, I hope to arrive at possible "solutions." To begin, I will explore the long-standing belief that it is possible to assess written work according to some absolute, reliable, and valid standard. By examining the relationship between the purposes and forms of evaluation, recognizing the factors which influence the evaluation of text written by students and read by teachers, and showing that evaluation cannot be all things to all purposes, I plan to show that this belief is a myth.

Another major issue is that evaluation of student work is very much a function of a teacher's perception of his or her professional role and purpose. This, in turn, is directly linked to one's belief in how meaning is made from and with text. Some teachers enact their role supportively and see textual meaning as open to negotiation between text, reader, and writer. For other teachers, their professional role is an authoritative and judging one, and textual meaning is largely fixed. Conflict between the facilitative and judging roles exists, and teachers who undertake the New Literacy find themselves struggling to resolve this tension. The fourth section of my paper will explore ways in which evaluation can support New Literacy models of reading and writing process while appeasing public demands for quantified measures of student writing ability. Finally, I will reflect on the politics of assessment and the challenges teachers face from a public demanding accountability in the form of reductive grades.

**The Myth of Reliable Marking Standards**

Myth can be defined in two opposing manners. One definition states that myth is "a traditional story or historical event which serves to explain a practice, belief, or phenomenon" (*Webster's New Collegiate Dictionary,* p. 754). Thus myth becomes "truth," not through empirical verification, but because it validates a psychological need to explain the inexplicable. But myth can also be "an ill-founded belief held uncritically, especially by an interested group" (*Webster's New Collegiate Dictionary,* p. 754). So, depending upon circumstances, myth can be either a truth or a falsehood.

I believe that "myth" appropriately describes attitudes held toward evaluation. Some individuals and groups believe that a number or letter is a valid description of student ability (Sarick, 1991, p. A1) and that evaluative measures can help to establish and verify national standards of student achievement. Whether the assessment is in the form of writing samples or objective multiple-choice tests of some of the skills writers must use in the construction of text, they believe that an evaluation yields an unassailable judgment of student ability and achievement, as measured against some predetermined standard. Literacy is "the ability to perform at a certain level on a standardized test and which asks education for preparation and practice in that ability" (Willinsky, 1990, p. 8). Others believe that most evaluative measures, especially those used in national or provincial testing programs, describe only single instances of student performance at specific tasks, and nothing more.

Ideological conflict exists between those subscribing to these opposing views of evaluation; each group sees the other's perspective as an ill-founded, uncritical belief. I intend to show that faith in an absolutely valid, reliable, predetermined standard of student ability or achievement in writing is the ill-founded belief, held uncritically by those with interests which run counter to the New Literacy.

Although the products of New Literacy learning are not always written text, I will focus on the evaluation of writing. It may be true that in all types of English curricula inordinate influence is placed upon written work (Chater, 1984), but evaluating writing, in its many forms, is the problem occupying classroom practitioners. From the outset, two fundamental oppositions underpin the problem of evaluation: the need for

evaluation which fosters a student's growth as a writer and the need for evaluation which serves such public or institutional agendas as assigning marks to an individual student, determining student promotion or retention, and obtaining a general measure of student literacy abilities across a large geographic area. Pauline Chater (1984, p. 18) describes the former type of assessment as *idiographic* (assessing the individual in his or her own right) and the latter as *nomothetic* (assessing students in comparison with others of the same age, grade, or level). She contends, rightly I think, that presently much assessment is nomothetic, whereas the nature of New Literacy tends toward the idiographic.

If evaluation is to be meaningful at all, everyone involved in the process--teachers, administrators, students, and parents--should know why and how student writing is and can be evaluated. Present evaluation policies of CCTE, NCTE, and the Manitoba Teachers' Society all support this position (CCTE, 1985; Manitoba Teachers' Society, 1989; NCTE, 1990), although there is a gap between the expectations behind these policies and public understanding of evaluation processes and measures. Student writing can be evaluated for a number of purposes, and it is important that purpose be determined from the outset in order that appropriate measures and procedures for evaluating student work be chosen. As Cooper and Odell (1977, p. ix) concisely delineate, three major uses of writing evaluation exist: *administrative, instructional,* and *evaluation/research.*

Evaluation may be undertaken for *administrative* purposes in order that student grades be assigned or predicted and that students be placed, tracked, or exempted from English courses. This is the expectation that the public most commonly has of an

English teacher's role as an evaluator. Such evaluation is also frequently undertaken by secondary or postsecondary institutions to determine placement of students in specialized programs. For example, Grade 9 students in the Winnipeg School Division No. 1 wishing entry into the Preparatory Year of the IB program must write a short diagnostic essay in addition to writing a version of CTBS.

"Exit" or "entrance" exams are becoming more common for students in their final year of high school; in Alberta, Grade 12 students wishing to earn a diploma must pass a provincial exam having both a multiple-choice and writing sample assessment (Willinsky & Bobie, 1986). Within the next few years, Manitoba students will be writing exams for similar purposes. Universities frequently use writing sample assessments to determine an entering student's need for remediation (Freedman & Robinson, 1982); they will also waive the necessity of a student's taking compulsory composition courses if he or she has attained high standing in a senior-year English course (University of Winnipeg, 1991, p. A-12).

Evaluation can also serve *instructional* purposes. Such evaluation is frequently described as formative evaluation, in contrast with summative measures and procedures, which serve the administrative purposes described above. Formative evaluation can provide both the teacher and the student with initial diagnosis of a student writer's problems, strengths, and potential. Moreover, formative evaluation or assessment is usually ongoing throughout the progress of a course, providing guidance and focused response to students as instruction continues. This is the perception that English teachers often have of their role as an evaluator, particularly in a New Literacy

classroom. However, the expectations, measures, and procedures of formative evaluation are usually at cross-purposes with those of administrative evaluation.

Finally, evaluative measures and procedures can be used for *evaluation* or *research* purposes. Determining the effectiveness of a writing program or curricular changes, measuring growth of individual or groups of student writers undertaking particular instructional treatment, and evaluating to accumulate data for any type of research project are all examples of evaluation for this third purpose.

Clearly, very different purposes exist. Regrettably, those involved or affected by the process of evaluation are not always aware of these differences, and this lack of awareness can lead to a misunderstanding of what test results actually represent. Whether in classroom instruction or in wide-scale programs, objective tests (usually tests having multiple-choice items with preselected answers) are misused for the summative evaluation of a student's abilities as a reader or writer. Tests of such writing subskills as punctuation, spelling, word usage, and grammatical correctness measure only the decontextualized knowledge of editorial skills, or mastery of certain writing conventions (Charney, 1984; Chater, 1984; Cooper, 1981; Cooper & Odell, 1977). While they can be used for purposes of prediction or for certain types of criterion measures in research studies (Diederich, 1974; Cooper & Odell, 1977), such tests are far too often used for administrative purposes and are totally inappropriate for the formative evaluation of student growth fundamental to a New Literacy approach. Their chief virtue lies in the comparatively low cost with which they can be administered and scored. Unfortunately,

in times of severe economic restraints, such a virtue can be elevated out of proportion to its value.

Throughout this paper, "evaluation of student writing" refers to the evaluation of actual writing samples, preferably those obtained in the course of classroom instruction. The findings of the literature I have examined for this paper are based upon student writing generated during the course of classroom composition instruction at the secondary and postsecondary levels as well as that generated for specific research studies. The classroom writing samples provide a real context for the evaluation problems faced by both teachers and students, whereas the purpose of data collection in many of the research studies was to examine problems which might lead to improvement of classroom instruction in writing, particularly from a New Literacy perspective.

Given the varied purposes which assessment can serve, what purpose underlies evaluation of student writing by teachers who adopt a New Literacy approach? Recall the second part of Willinsky's (1990) definition of the New Literacy: "a social process with language that can from the very beginning extend the students' range of meaning and connection" (p. 8). I believe that the key word, in this part of the definition, is *extend.* It implies growth, development, moving beyond the point at which a process began. Much institutionally mandated evaluation seeks to sum up, in one letter or number, a student's apparent achievement, rather than charting the changes, development, and extensions of ability that a student has undergone in order to reach the point at which he or she is evaluated. In the fourth section of this paper, I will return to this issue and discuss the ways in which teachers can address the apparent

process-product dichotomy. At this point, however, I want to focus on the problems arising for both the teacher-evaluator and the student writer when evaluation for administrative purposes must be undertaken.

When I began researching literature on the evaluation of student writing, I was struck by the comparative lack of information available on assessing secondary student writing in modes or forms other than the argumentative essay. The essay is the overwhelming choice both for wide-scale assessments of high school students in academic streams (Willinsky & Bobie, 1986) and for university placement and proficiency tests (Freedman & Robinson, 1982; Hoetker, 1982). However, some wide-scale programs, such as the NAEP assessments of writing, provide students with narrative and informative writing tasks (Applebee et al., 1990) in addition to a persuasive writing task. Research studies also favor the personal or persuasive essay as the preferred mode and form of student writing samples.

This lack of information on evaluating writing in other forms and modes is one of the major difficulties faced by teachers justifying New Literacy initiatives. The body of research which might conclusively demonstrate its worth simply does not exist. In fairness, it can also be said that insufficient research exists to refute its value (Willinsky, 1990). This section of my paper focuses on the impossibility of attaining an absolute standard of writing assessment, and if such a standard cannot be attained for the argumentative essay, I find it highly unlikely that such a standard can be attained for the formative evaluation of the richly diverse writing which comes out of New Literacy classrooms. Although a problem, I see this as one of the many challenges the New

Literacy faces in proving its value to those skeptical of its worth.

Even if the essay were the only discourse form in which students wrote, rendering a reliable, valid, and absolutely standardized evaluation of writing is impossible. Some of the impossibility arises out of three problems faced by the student writer: (1) the situation of writing for an evaluation, (2) the problem of interpreting the demands of a topic or prompt, and (3) the problem of knowing the demands or conventions of a given discourse mode.

Knowing that his or her paper will be read and evaluated cannot help but influence a student's writing. Teachers may intrinsically value reading and writing and may work very hard to develop in their students a similar regard. However, most academic reading and writing takes place in a climate that is highly evaluative, "in which grades are exchanged for performance" (Nelson, 1990, p. 365). Rather than investing themselves personally in their writing, students often attempt to find the performance formula which they believe will yield the grade that they seek.

In *Response to Student Writing* (1987), Sarah Warshauer Freedman relates the story of Jody, a first-year college student who chronicles her writing experience and her teachers' attitudes toward her writing. For Jody, the temperature of the evaluative climate warms up or cools down depending upon whether or not she writes what she believes the teacher wants to read. Grades and a teacher's evaluative responses can powerfully influence a student writer's text; for the reward of a grade, students stop writing in order to develop and explore their own rhetorical needs and "just write for other people" (Freedman, 1987, p. 3). Writing becomes a matter of following rules made

by others, rather than trying to discover for oneself the rules which work in one's own writing.

Furthermore, in her study of teacher response to student writing, Freedman (1987) discovered that although students found formative evaluation, in the form of teacher commentary, to be helpful during the drafting process, they wanted grades on their final versions, in order to gauge how correctly they had interpreted the "'rules' of good writing" (p. 78). The following student's comment, familiar to English teachers everywhere, reflects the influence of grading on student text:

> I think my teacher helps me a lot, because if it's not done the right way you'll get a bad grade. And getting a bad grade really hurts, so the next time you'll do it the correct way, so you can see some improvement, and better grades. (Freedman, 1987, pp. 78-80)

No student writer, whether in grade school or grad school, wants a bad mark. So, the writer seeks to align his or her text with the perceived audience, the teacher-evaluator. Audience awareness is an important rhetorical element in any piece of writing, but overconcern with how that audience will evaluate a text leads students to suppress certain elements, enhance others, employ strategies with which they are uncomfortable, and support interpretations for which they have only superficial belief. James Marshall's (1987) study of the effects of writing on student understanding of literary text contained interviews with students who believed that certain rules prevailed

in the writing of argumentative essays in their Old Literacy classroom. Indeed, the most successful student in the study clearly understood how to apply the rules in order to shape his writing into a text for which he would be rewarded with an A.

The teachers in Freedman's study were, to varying degrees, advocates for New Literacy teaching of reading and writing; the teacher who was the focus of Marshall's study was not. However, even when teachers use process methodology, students will sometimes subvert and circumvent the learning process intended by an instructor. Jennie Nelson's (1990) study examined writing assignments in three very different disciplines; yet despite the instructors' strong expectations that students work hard at defining and conceptualizing their writing tasks, a number of students used alternative strategies, circumventing the processes instructors expected them to use. Why did they do so? They felt confident that they and their writing would be rewarded with good to adequate grades, even when they did not use process writing strategies.

Classroom teachers are familiar with the student who refuses to "buy into" process writing, preferring the one-shot effort for which he or she has usually been adequately rewarded. Even more familiar is the student who, "in weigh[ing] the risks in choosing a particular approach or answer" (Nelson, 1990, p. 365), decides not to risk his or her potential grade by trying a new approach or suggesting an unusual answer. Obviously this is not the way to extend one's skills as a writer, but given the often chilly nature of the evaluative climate, students may decide that the risk-taking inherent in New Literacy may not be worth the trouble.

All writers take risks when they write. They risk being misunderstood, disdained, or simply ignored. Knowing that their work will be graded, student writers face a riskier situation. Even in classroom situations where teachers and students work together as a highly supportive community, the spectre of the bad grade can haunt a confident but grade-conscious student.

Topic and mode of a writing assignment also affect a student writer, and by extension, the evaluation he or she receives. In a New Literacy environment, virtually everything and anything can be the stimulus for writing, and virtually any type of writing is produced (Atwell, 1987; Willinsky, 1990). However, administrative purposes--placement exams, wide-scale assessments, year-end examinations, and sometimes, school examination policies--usually mandate that students write on topics or prompts and in modes preselected by the team which creates the evaluative instrument. Classroom teachers, particularly those who have served on holistic marking committees, can attest to the difficulties that students can have in aligning their perceptions of a topic and discourse mode with their ability to produce them.

Should topics be wide open or narrowly defined? Should they open up the potential for a student's written response, both in form and content, or should they constrain response in order that raters can achieve a higher degree of reliability when they grade papers (Charney, 1984, p. 70)? In describing their experience with the testing of student writing proficiency at San Francisco State University, Freedman and Robinson (1982) provide some worthwhile, common sense observations on topic design, although

they carefully point out that the suggestions are intended only for expository writing topics.

Topics must be "accessible"; that is, they must draw upon background knowledge and concerns which are likely to be common to the students writing the test. An example of a topic containing sufficient range in these areas might be to ask students to write about "any kind of gripe, major or minor, or any kind of character weakness, silly or significant" (Freedman & Robinson, 1982, p. 395). The topics must discourage clichéd or platitudinous responses; responses drawing heavily upon religious or political beliefs may contain language or concepts which are not the student's own (Freedman & Robinson, 1982, p. 396). Topics should not be of such depth that students cannot write about them in the time allowed, and they should elicit the mode which the evaluators intend; that is, the topic should guide the student to write exposition if that is what the testers expect. Freedman and Robinson also advocate extensive pretesting of topics to ensure that all of the above conditions are satisfied; for wide-scale assessments of any sort, this pretesting is likely to uncover potential problems.

Although a topic can create difficulty, discourse mode usually poses much more of a problem for the student writer (Nold & Freedman, 1977; Hoetker, 1982). The various discourse models described by Lloyd-Jones (1977) indicate the difficulty of neatly classifying pieces of writing into discrete and qualitatively different modes. Even when evaluators can agree on a discourse model as the basis both for topic/prompt design and for designing a scoring rubric, the difficulty is not yet resolved. Classroom teachers know from experience that students do not always produce the expected form of writing. A

student may be unable to write in the mode required or simply may not want to write in that mode. Secondary English teachers have all experienced the dismay of reading a plot summary when argument was expected.

Clearly, student writers are caught in a complex web of factors affecting their writing, particularly in a test situation. The situation of being evaluated, the topic about which they write, the mode in which they write--all of these influence writers and their writing. But what influences the readers of these texts, the teachers who will evaluate them? How reliable and valid are the judgments of teacher-evaluators?

The concepts of reliability and validity are crucial to any discussion of evaluation, and in the assessment of writing, the search for measures both reliable and valid has been ongoing and largely unsuccessful. Davida Charney's (1984) critical overview of the use of holistic scoring to evaluate writing defines these two concepts concisely:

> A *reliable* measurement is capable of replication under equivalent conditions. So, a reliable method of assessing writing ability would yield a consistent judgment of a student's abilities if applied again, all else being equal. A *valid* measurement assesses what it claims to assess. So, a valid writing assessment would be sensitive to a writer's "true" abilities. (p. 65)

I would argue that those who cry loudest for an absolute standard of quality in a writing assessment are also the least aware of the difficulties of satisfying the criteria of both reliability and validity.

Even holistic scoring, described as a "quick, impressionistic qualitative procedure for sorting or ranking samples of writing" (Charney, 1984, p. 67), has its limitations, and is best used for administrative evaluations such as placement exams or wide-scale assessments. It is also used in research studies, although in a research study the degree of inter-rater reliability must be quite high (usually somewhere between .85 and .90). Various scales can be used (Cooper, 1977), but most fall into one of two types: (1) general impression marking, in which the grade is assigned on the basis of a total impression the writing makes upon the reader, and (2) marking based upon a scoring guide or scale listing specific discourse features which the reader must keep in mind while reading and ranking the papers.

Holistic marking is not the solution to the quest for the absolute standard. Readers are trained to make judgments, either according to criteria which they have arrived at or against a set of criteria which have been formulated in advance of a session, but discrepancy still arises. Readers are thus cautioned to monitor their own marking in order that they maintain reliability, and in cases where two different raters arrive at highly discrepant marks, another opinion will be sought from another reader. So, holistic marking takes place under fairly controlled conditions, yet normally some degree of inconsistency is seen. What causes discrepancies to arise between raters? Or, to put the question another way, why can trained, competent readers not arrive at exactly the same evaluation of the same piece of writing?

Paul Diederich's *Measuring Growth in English* (1974) contains an account of his 1961 study, *Factors in Judgment of Writing Ability,* and gives some insight into why total consistency amongst readers, and by extension, an absolute standard of writing assessment, cannot be achieved. In his study, sixty different readers from six different occupational fields read, ranked, and commented upon three hundred papers written by first-year college students. Readers were to rank the papers in order of merit, from 1 to 9, and were asked to write comments describing the features of the writing influencing their evaluation.

The resulting inconsistencies in the ranking are surprising only to those who have not participated in group or holistic marking sessions: no essay received less than five different grades, and 101 essays received every possible grade from 1 to 9. The comments written by the readers clustered into five major groups, indicating major features of writing to which they reacted: ideas, mechanics, organization, wording or phrasing, and finally, "flavor," which might also be described as style, or possibly, as voice.

What leads to such incredible inconsistency? Even though the participants in Diederich's study were not all specialists in language arts or English, various studies of evaluation of student writing by English teachers indicate a similar variety of biases. Knowing the identity of the writer, agreement or disagreement with a particular ideological slant, or being highly influenced by one of the five factors listed above can and does affect a reader's evaluation of a text (Diederich, 1974; Freedman, 1979). Handwriting can also adversely and unfairly influence a student's grade (Chater, 1984;

Charney, 1984). Nold and Freedman (1977) found that length of an essay correlates curiously as a predictor of a high grade: "it is more damning to write a short essay than elevating to write a long one" (p. 173). And even when teachers claim that certain features of writing are highly influential in their evaluation, in practice, what they really value is quite different. Winifred Hall Harris's (1977) study showed the very strong negative influence of mechanical error upon teacher evaluation of student writing, even when teachers claimed that they valued content and organization more highly. A paper deficient in mechanics or word usage but competent in content and organization was twice as likely to receive a lower rank than a paper competent in mechanics and usage but deficient in content and organization (Harris, 1977, p. 185).

Thus, even in the reasonably controlled conditions of holistic marking, there is no absolute standard; the scores obtained are relative only to the particular set of papers being evaluated for that particular writing task, the age and ability of the student writers, and the particular conditions under which the papers have been written. Raters must reach consensus and be trained to conform to criteria set by them or for them; they cannot use "an absolute standard of quality, perhaps that of published adult writing" (Cooper, 1977, p. 20), because such a standard simply does not exist.

If absolutely consistent evaluation of student writing cannot be obtained in controlled conditions, how can it be possible in the highly variable situation of day-to-day classroom life? Students and teachers, as writers and readers of text, are working amidst a variety of personal and social conditions which ultimately influence the evaluation of texts read and written. The expectation that teachers and students direct their efforts

toward meeting an absolutely reliable, valid, and standard assessment of writing achievement and ability must be recognized for what it is: an impossible and untenable position. Such expectations place teachers in the position of final judge of the quality of writing without acknowledging the fallibility of that judgment. It is a heavy responsibility, and I know that some teachers are quite happy to divest it to an outside agency, such as an IB or AP marking board.

Earlier, I stated that all involved or affected by the process of evaluation should know why and how students are evaluated. I also believe that they should know the limitations of evaluative measures and the validity of scores obtained from such instruments. However, for teachers, administrators, and policy-makers to share such knowledge would mean that they also must share the power that comes with this knowledge. The New Literacy seeks to share that power, to shift control and responsibility to the student. Yet doing so entails a crucial recognition of both the limits of a teacher's authority and the potential for student empowerment.

**Coach or Judge: Which Hat Fits?**

In *Readings and Feelings: An Introduction to Subjective Criticism,* David Bleich (1975) comments that "most new ideas in teaching try in some way to get around the need for a short significating evaluation" (p. 105). Still, Bleich acknowledges that the need remains, usually because the administrative purposes of evaluation have to be served. He also acknowledges the difficulty of evaluating literary response and thus offers a variety of suggestions as to how an instructor might implement such responses in a New Literacy

classroom. I have already pointed out that highly subjective factors can influence teacher response to student writing; Bleich admits not only to this subjectivity, but to possible "abuse of this subjective exercise of authority" (p. 108). In making this point, I believe that he addresses a key issue in evaluation: the question of power and authority, who has it and who is willing to share it.

To no small degree, grading is the exercise of teacher power over the student. In the previous section of this paper, I have shown that student awareness of this power can have telling effects upon their writing, such that they write not for personal satisfaction or the extension of their abilities, but for the grade which rewards their efforts. Teachers often wield this power with the best of intentions, believing that high expectations will necessarily lead to high levels of effort and performance. Similarly, students socialized by the values of school frequently equate the difficulty of obtaining a high grade with the value of a course.

It is often said that teachers teach as they have been taught; by extension, English teachers grade as they have been graded. If, as students, their papers were copiously annotated in red ink, every error highlighted for scrutiny as well as for justification of a particular grade, they may become "forensic graders" as teachers (Baumlin & Baumlin, 1989, p. 176), seeking out and punishing errors. With this cast of mind, the teacher-evaluator sits as judge of the student's writing, and the judgment rendered is final. In such an environment, "the student's text, simultaneously the scene and perpetration of its own crime, becomes an object or event frozen both in time and in its present shape"; the teacher shows "little or no interest in the growth of this text or the possibility of

future change--that is, textually speaking, in the possibility of *revision*" (Baumlin & Baumlin, 1989, p. 177).

From a New Literacy perspective, it is easy to condemn the forensic grader (if one is not blushing with shame at the recognition of a former self). After all, such a teacher is either ignorant of the move toward process pedagogy or is simply ignoring these changes, actively or passively. Still, I believe that we must recognize how difficult it is for teachers to surrender the authority vested in them by tradition and school culture. Undertaking New Literacy's shift of control from teacher to student involves a major ideological change, a conversion experience. My readings of other practitioners of New Literacy lead me to speculate that a teacher must reach a fairly high level of self-assurance in his or her abilities before the shift is accomplished. The power relationships implicit in reading and evaluating student writing undergo major changes in a New Literacy environment, and for many teachers, great personal confidence must develop in order that offering students control is not threatening.

Fundamental to a change in the power relationship is the teacher's personal recognition that he or she is uncomfortable with the role of evaluative judge. Elizabeth Flynn (1989) advances the interesting claim that her discomfort with the judging stance, and subsequent conversion to a more supportive, sympathetic stance, was gender-based. She identifies her move toward a process-based pedagogy as a feminine stance, one that values empathy over detachment and objectification and caring about student development over punishment for crimes of composition. James Corder's (1989) perceptions of how he has been evaluated (by administrators, by teacher evaluation

questionnaires, by student responses to his own writing) sensitized him to the difficulty of rendering a judgment on a student's work, both on individual pieces of writing and throughout the course of a semester or term. In these two very different examples, both teachers manifest awareness of the inequity of power in the teacher-student relationship, the manner in which it is enhanced when teachers are vested with the power to evaluate a student's work, and the difficulty of equalizing the balance of power.

Role strain may develop, not only from personal discomfort with an inappropriate role, but also from the cognitive shift which takes place when both theory and experience merge in action and lead to a reconceptualization of one's professional practice. In the first section of my paper, I wrote of the ways in which personal reflection, awareness of new pedagogical theory, and changes in my own teaching practices caused me to think differently about evaluation of student writing. Theories of evaluation are directly linked to questions of how meaning is created in text (Gere, 1980; White, 1984; Lawson & Ryan, 1989; Winterowd, 1989). For this reason, teachers of the New Literacy, using process pedagogy as their model of writing instruction and reader-response criticism as their dominant approach to literature instruction (Willinsky, 1990), necessarily find that methodological change entails a changed role as a teacher.

Thus, for teachers of the New Literacy, evaluation is truly formative; it is ongoing response to the student's texts while they are in the process of being formed and composed, rather than a final judgment of the text after it has been completed. Teacher response to a student's writing, either through a conference or written commentary, serves to motivate the student to keep the process of drafting and revision going

(Sommers, 1982). Just as writing should not be a "one-shot" effort, neither should evaluation be a "one-time" service provided by the teacher when the student has submitted the final copy of a text. Moreover, I believe that such response helps to sustain the student's involvement with the text from beginning to end, and in doing so, serves the fundamental purpose of shifting control of literacy from teacher to student.

Previously, I listed three main purposes which evaluation can serve. In all three of these purposes, however, assessment is usually initiated by someone other than the student, and often for reasons which will benefit someone other than the student. If students are to gain control of their literacy, then they must be participants in the process of evaluation and not mere passive recipients of commentary and grades. Control comes from learning to assess and evaluate one's own writing; it is not developed if one is totally dependent on someone else's judgment. However, for students to gain that control, teachers must give up some of theirs. Just as no form of summative evaluation of writing is ever absolutely reliable, formative evaluation poses its own set of problems for the New Literacy teacher.

Composing models of reading and writing (Tierney & Pearson, 1983; White, 1984; Crowley, 1989) imply that the teacher who reads and responds to a text is constructing meaning and "rewriting" the text through that process of meaning construction. In fact, the commentary which teachers write on student drafts-in-progress is usually intended to give the student writer a sense of the meaning which the teacher has constructed from his or her reading of the paper and to sensitize the student writer to questions and issues which may not have occurred to him or her (Sommers, 1982). So far, so good.

However, our teacherly intentions of serving as a guide and a coach of a student's growth and extension as a writer can be subverted by the nature of the power relationship which traditionally exists between teachers and students.

Quite simply, it is all too easy for teachers to read their version of the text over top of the student's, and because teachers are in a position of authority over the student, "the student will try to do whatever he is told to do by a teacher" (Crowley, 1989, p. 107). And why not? By virtue of their professional expertise, teachers are supposed to "know better," and students should follow the advice given. If students do follow the advice given, whose text is it now? Ours or theirs? Whose work is being evaluated? Furthermore, how do teachers react to students who choose to disregard our advice, however well-intentioned it may be? When serving as co-author and editorial assistant, how do teachers avoid appropriating the student's text and exercising the power of teacher authority over it?

Nancy Sommers's 1982 research on teachers' evaluative response to student writing revealed that "students make the changes the teacher wants rather than those that the student perceives are necessary, since the teachers' concerns imposed on the text create the reasons for the subsequent changes" (pp. 149-50). Sommers's collaborators in the research project, Brannon & Knoblauch (1982), make a similar case for the problem of appropriation of the student text by the teacher-reader, but from a slightly different angle. Rather than focusing on the nature of teacher commentary in the reading of student writing, they focused more directly on the issue of a writer's authority and ownership of text.

A writer's renown and authority enable readers to tolerate writing which causes difficulty and confusion (Brannon & Knoblauch, 1982, pp. 157-58); we do not expect that Eliot, Joyce, or Beckett rewrite their texts so that they are more accessible to us. However, teachers are not as tolerant of ambiguity in student writing; not only do teachers

> view themselves as the authorities, intellectually maturer, rhetorically more experienced, technically more expert than their apprentice writers, . . . [but] in classroom writing situations, the reader assumes primary control of the choices that writers make, feeling perfectly free to "correct" those choices any time an apprentice deviates from the teacher-reader's conception of what the developing text "ought" to look like or "ought" to be doing. (Brannon & Knoblauch, 1982, p. 158)

Now, I believe it must be acknowledged that, in many cases, teachers do know more than their students, do have more experience and expertise as writers, and can offer their students real insight into possible directions in which a text can be taken. But it takes considerable restraint to keep this intellectual authority in check. For those of us who have been rewarded for demonstrations of expertise, suppressing that side of ourselves is no easy feat. It is much easier to take the student in hand and say, "This is how you should revise this piece," than to work at eliciting possible approaches from him or her. Still, if we sincerely expect to enact New Literacy in our classroom, we must

make the effort and find ways to preserve the student writer's authority.

Renegotiating the balance of power and surrendering authority also demands a new perception of the function of error in student writing. It is all too easy to perceive error as a failure to learn and punish the student for perceived ineptitude, rather than see error as an unsuccessful attempt at trying a new skill or at extending range of performance. Teachers in New Literacy environments often see themselves in a "coaching" role, and it is useful, at this point, to examine a coach's attitude toward training and skill development.

Coaches expect that their trainees will make technical or tactical errors in the course of training. When the trainee errs, the coach's role is to help him or her see those errors in terms of performance goals and then to formulate a strategy to improve the next attempt. When the trainee is successful, the coach's role is to help the trainee analyze what he or she did right in order to attain the desired performance goal, thus sustaining and improving the technique which led to success.

Both the coach and trainee are actively working together at the goal of improved performance, a model of combined action which underlies New Literacy instruction. While the coach probably has greater experience than the trainee, he or she may or may not have more innate talent, ability, or motivation, and it is on this last quality which the coach must sometimes draw the most. So it is with teachers and student writers; teachers must find ways of "nudging" the student, to use Atwell's (1987) term, toward the motivation to take risks in their writing, even though risks necessarily entail error for the novice. Although a teacher may have the authority which comes from experience and

expertise, both teacher and student must work together to maximize the opportunity of a learning situation.

Mina Shaughnessy's (1977) work with basic writers revolutionized perceptions of error. Students do not make errors out of unwillingness to learn or out of spite, although, as Nancy Sommers (1982) pointed out, there can be remarkable meanness in teacher response to student writing (p. 149). Errors are a natural consequence of students' struggles to achieve their writing goals.

I offer, as an example, a thank-you note just received from the seven-year-old daughter of my friend. It reads:

> Dear Joanne and Michael
> Thank you for the Goebel [a porcelain figurine]
> and for comeing to my first communion The Goebel
> look's nice in my room
> Love
> Megan

Megan has been reading and printing for two years, but her thank-you note indicates that she still has problems forming and spacing letters evenly. Two common words are misspelled and punctuation is absent, although she did correctly spell a foreign name (Goebel) and a polysyllabic word (communion). Ten years ago, I would probably have railed at the ineptitude of her teachers and marvelled that her mother would even allow

such a note to be sent. While I must confess that I am still jarred visually by the errors, I am now much more appreciative of her efforts to write her own message, mistakes and all.

At the same time, those mistakes remind me of the ways in which student texts really are different from many other forms of writing. They contain features "such as spelling errors, structural defects, and solecisms [which] make special demands upon a reader" (Lawson & Ryan, 1989, p. ix). The exhaustion which many of us experience as we read student writing (Ede, 1989) is undoubtedly due to the extraordinary effort we must give to the task of making meaning, particularly when the text causes us interference. At the same time, many of us have, in the course of our studies of literature, persevered with T. S. Eliot or James Joyce, even though their texts also contain gaps, fragments, and syntactic convolutions. When does "error" become art? I suggest that it is when we accept the writer's authority to make errors in pursuit of extending the "range of meaning and connection" (Willinsky, 1990, p. 8).

Still, I am not naive about the limits--my own as well as others'--to acceptance of error. It may be true that English teachers are "hyperliterate" (Murray, 1989, p. 81) and overly sensitive to error, but when the cry goes out that students cannot write, this opinion is often based on the examination of surface errors in written discourse. The type of error I am willing to accept from a seven-year-old is not the same type of error I expect to see in the writing of a seventeen-year-old, but nevertheless do. Error can be an indicator of growth, but teachers also have to accept that there may be limits to growth. And the growth of the writer may not lead to an improved text.

This is difficult for both students and teachers to accept. Traditionally, the student trusts the teacher to diagnose the problem in his or her text and offer suggestions for solving it. When the suggestions do not lead to immediate and obvious improvement, the student may begin to doubt the expertise or authority of the teacher, although there may be good reason why the suggestion did not work. Whatever the case, both teacher and student experience frustration. The teacher is frustrated because his or her best efforts at aiding the student have not yielded recognizable change in the student's work. The student is frustrated because, despite having taken the teacher's advice, he or she has not achieved the desired writing goal: an improved text, and possibly, an improved grade.

Both William Irmscher (1979) and Knoblauch and Brannon (1984) examine the problem of evaluation and the "myth" of improvement. They both view error as evidence of lack of writer control over structures or text features. But when it comes to actual evaluation of the text, they take somewhat different directions. Irmscher states, rightly I think, that "students expect progress to be registered continuously by gradations, although the teacher may soon reach a ceiling grade" (p. 153). However, Irmscher fails to address the question of how the teacher maintains credibility when the ceiling grade is reached. Knoblauch and Brannon argue for patience: "symptoms of growth--the willingness to take risks, to profit from advice, to revise, to make recommendations to others--may appear quickly, even if improved performance takes longer" (p. 169). Yet the time constraints of present public school teaching and learning conditions frequently work

against development of this type of attitudinal change.

Clearly, in a New Literacy approach to evaluating writing, a nonpunitive attitude toward error is demanded, along with patience on the part of both teacher and student. The teacher must find the patience to hold back on exerting his or her authority over the student's text; the student must find the patience to accept that growth is often a slow and painful process, with error as a necessary part. The teacher's authority to judge must move toward the willingness to facilitate growth. If we relinquish authority over our students' texts and accept that making meaning from text is an ongoing process, then perhaps we can see grades only as the response to a paper at the time we read it, and not as the ultimate judgment of that piece of writing.

**Alternatives in Assessment**

In talking with other teachers, I have often found that they describe their students' experiences of the New Literacy with conviction, enthusiasm, and genuine excitement. When the talk turns to evaluation, though, enthusiasm fades and gives way to profound doubt. Teachers expect, but do not always enjoy, having to justify their work to others: to colleagues, who may not be enamored with these "newfangled" innovations; to administrators, sensitive to public and district pressures for "standards"; to parents, concerned that their children emerge from school with a solid grounding in the "basics" of literacy; and finally, to students, worried that they really will not be adequately prepared for the rigors of postsecondary studies.

I believe that this crisis of confidence stems largely from a lack of generally accepted alternative models of assessment which support New Literacy. Two decades of New Literacy theory and practice (Willinsky, 1990) have not yet yielded the body of research which might foster acceptance. Furthermore, even when research validates classroom experiences of the unreliability of assessment, the love-hate relationship teachers have with researchers and credentialed experts is a virtual guarantee of their continuing to evaluate as they have.

However, if teachers are really interested in taking on the challenge of the New Literacy, of sharing with their students the responsibility for developing literacy abilities, they must do what all of us do when we write and read: take a risk. Perhaps reassurance can be found in Mina Shaughnessy's (1977) notion of error as an inherent aspect of growth as a learner, for New Literacy teachers are self-reflective students as well as teachers. Indeed, one of the cornerstones of such works as *Encountering Student Texts* (1989) or *In the Middle* (1987) is the authors' unabashed use of "I," continually reminding the reader that the subject is theory into practice, reperceived through highly personal experiences of both.

Taking a risk entails the possibility of making an error, but that is the only way in which new knowledge is formulated. The New Literacy entails new approaches to assessment; one hopes that, ultimately, new research and increased information on classroom practice will validate these new approaches, finally putting old myths to rest. Teaching, however, is a conservative profession, "where nothing is known unless the 'right' people know it" (White, 1984, pp. 194-95). "Rightness" is an ideological

determination, and the "right" method or the "right" people are the "wrong" method or "wrong" people to those skeptical of or hostile to the New Literacy. At this point, perhaps the best that any of us can do is to try a variety of the methods described below, keep careful notes, and share our experiences with the entire constituency for whom evaluation is a crucial issue.

Central to the New Literacy is to focus on the process of literacy as much as its products. In the statement of rationale for the Manitoba English Language Arts 9-12 Curriculum, "evaluation of process as well as product" (Manitoba Education, 1987, p. 4) is the first element listed. Given that evaluation of writing products can be unreliable, teachers often ask, "How is writing 'process' to be evaluated in quantifiable terms?" When easy answers are not available (and easy answers do not exist), teachers are then led to create marking schemes for "process." For instance, in a given assignment, students might obtain five marks for a web, outline, list, or other evidence of prewriting; five marks for an initial draft; and ten marks for a final draft.

I believe that such an approach misdirects the main intention of evaluation in a New Literacy classroom: "formative evaluation for the purpose of identifying and responding to students' strengths and weaknesses" as well as "developing students' evaluative skills so that they can take increasing responsibility for their own learning" (Manitoba Education, 1987, p. 4). It is impossible to quantify the diverse elements which work together to produce a piece of writing; the best that can be achieved is a relative judgment of how well a writer used whatever strategies and resources he or she had at his or her disposal. This is not, I realize, the answer that most classroom teachers want

to hear, but it is the answer at which many New Literacy teachers have arrived (Bleich, 1975; Atwell, 1987; Probst, 1988; Gilbert, 1990).

An increasingly popular means of providing evidence of both the process and products of a student's writing is the writing folder, file, or portfolio. Typically, portfolios contain student writing in a variety of genres, modes, and forms; individual pieces may be graded, or pieces may be selected as representing the body of a student's work and a grade assigned on the basis of those representative pieces. Folders may be collections of a single year's work, or they can be cumulative, with pieces from previous years retained so that both student and teacher can trace the writer's development.

Similarly, there are two main approaches to folder management. In the first, writing folders are like artists' or designers' portfolios, containing only representative works in certain genres or media or works judged by the writer to be among his or her best. Using such an approach, students may keep working drafts in another form or in another location (journals, paper copies, or computer files). The portfolio is thus a selection, rather than a collection, of a student's entire work for the year.

The alternative approach is thoroughly described by Nancie Atwell (1987) in her mini-lesson on writing workshops (p. 83). She tells her Grade 8 students that "you're creating a history of yourself as a writer this year" (p. 83), and as history only becomes tidy and systematized when it is sorted into some type of organized pattern, it is a decidedly eclectic collection. Students are expected to use their folder for works-in-progress, for keeping note of future projects they might undertake, and for keeping their own record of the skills they have learned in the course of the year (p. 85). Atwell quite

openly acknowledges the tension between fulfilling an administrative requirement for a grade four times a year, her own need to honor the nonlinear progress of writing growth, and her students' need to risk and experiment as they write (p. 114).

In addition to surveying the material in the folders, Atwell holds an evaluation conference with each student, asking them about their perception of the process of effective writing, about their assessment of their best pieces of writing, and about future writing goals, plans, and projects (pp. 114-16). Because report cards mandate them, she uses a scale of letter grades, and in subsequent evaluation periods after the first term, she "bases a writer's grade on progress made toward the individual goals established in the evaluation conference" (p. 119). Students who completely accomplish their goals receive A's, good or adequate progress is rewarded with B's, adequate or fair work receives C's, and so on. Hers is an interesting compromise, and although administrative demands mandate the use of conventional letter grades, her focus is on formative assessment and on developing student awareness of the ability to self-evaluate.

Maintaining a balance between teaching curriculum-mandated composition forms and providing secondary students with the opportunity to choose their own topics and forms can also be accommodated through portfolio assessment procedures. Mike Gilbert (1990) works at this balance through the use of a two-tier portfolio grading system. He provides an individual grade and supportive response for individual papers, but also evaluates the student's entire collection of material. Students then receive an overall portfolio mark which reflects half of their grade for that marking period. In this way, no single piece of writing, whether highly effective or totally disastrous, skews the student's

grade. Like Atwell, Gilbert recognizes the inevitability of grades and has developed this compromise in order to give the greater portion of his energies to responding to student work-in-progress.

Atwell and Gilbert directly address the conflict between administrative and instructional purposes in assessment and in trying to find means of building in recognition of process while surveying a variety of writing products. Robert Probst (1988) is not so sanguine in dealing with the process-product dichotomy. Although he provides checklists of criteria to consider when evaluating a student's response to literature, he gives little practical advice that can be readily implemented in a classroom. However, his reflections on an educational system which seems to prefer the "meaningless simplicity" of letter or numeric grades to the "meaningful complexity" (p. 224) of assessing "the ability to create, to imagine, to relate one thought to another, to organize, to reason, or to catch the nuances of English prose" (p. 221) certainly offers another perspective on the problem.

Actively involving students in the process of self-evaluation is fundamental to their development as independent writers and learners capable of judging when to seek out another opinion and when to listen to their own. Although the goal-setting which Atwell incorporates into her evaluation conferences might seem revolutionary to some, it is the first of three types of student-centered evaluative processes described by Mary Beaven (1977). Beaven's awareness of the influence which years of teacher dependency can have upon students (p. 153) leads her to suggest beginning with individual goal-setting as the first of her three strategies. In this approach, a great deal of teacher support is built in,

with the teacher diagnosing both the strengths and weaknesses of a paper and both the student and teacher choosing one writing problem as the goal toward which the student will work in the next assignment. The teacher's evaluation of the next assignment is then targeted to the goal which teacher and student have mutually agreed upon.

Self-evaluation can be the next step. In self-evaluation, rather than the teacher suggesting the goal toward which the student should work, the student responds to a list of self-assessing questions which enable him or her to focus on both the strengths and weaknesses of the paper, major structural concerns, minor mechanical problems, and the student's own perceptions of what he or she was trying to accomplish in the paper. This self-evaluative model can be extended to peer evaluation groups, although Beaven is quick to point out that the class in which it is to be used should have established a high level of trust and be well trained in group process. Moreover, the teacher using peer evaluation groups must be prepared for the fact that group process takes time; "groups that function well tend to spend half their time on process and half on task" (Beaven, 1977, p. 152). Demands for "coverage" of a particular block of material can work against both the teacher's and students' best intentions of effective group process.

The writing generated in New Literacy classrooms is not always the formal essay; in fact, the diversity of written products is often quite amazing to those of us who were taught and began teaching in Old Literacy classrooms. But how do teachers evaluate genres which would be classified as literature if they were written by a well-known author? How do teachers grade literature response logs or writing journals, forms of

writing whose discourse features defy many of the conventions of either narrative or expository text?

The evaluation of journals is a particularly thorny issue. Certainly journals are "the most idiosyncratic and variable" (Fulwiler, 1987, p. 7) of all writing assignments. Like Richard Beach (1989), I use assignment or literary response journals to foster reflection and response to discussion, readings, viewings, and writing. Whether through ongoing dialogue response or "nudging" (Atwell, 1987) commentary, I work at communicating very directly, at extending reach and connection with my students as another human being and not just as an English teacher.

I would prefer not to grade journals, but know that such a preference is, in present school culture, unrealistic. Senior high students, conditioned by years of schooling to accept payment in grades for all that they write, can remain unconvinced that an assignment is "serious" unless there is a mark value attached to it. I have often graded journals by using a Pass/Fail/Honors mark, with the grade open to student-teacher negotiation. At reporting periods, however, I am forced to convert Pass/Fail/Honors designations into a number. But if Nancie Atwell can rationalize her conduct, so can I.

David Bleich (1975) also would prefer not to grade at all (p. 105), but after acknowledging its necessity, he offers admittedly vague suggestions for evaluating responses to literature. He describes his assessments as being both quantitative and qualitative. His definition of "quantitative" is literal, as it is based upon the amount of work a student produces; the "qualitative" aspects are derived from a teacher's evaluation

of the seriousness of purpose behind the production of the work. How does one determine seriousness of purpose? Bleich says only that grading in a response criticism course comes to a decision "between what is adequate and what is excellent" (p. 109).

Lacking any other, Bleich's criteria can be applied to the even more difficult task of evaluating student-produced literature. Maintaining portfolio collections and using evaluation conferences of the type described by Atwell seem to be two of the most workable approaches to assessing student writing in discourse modes with which most teachers are unfamiliar as writers. Even Atwell does not give much advice specific to this issue, though individual goal-setting is behind much of her general evaluative intent, and it is certainly applicable to the evaluation of what is often termed "creative" writing. It is a curious irony of the New Literacy that student-written literature is valued to a degree previously unseen, but methods of evaluating that literature are still in the development stage.

However much teachers of the New Literacy are challenged by the instructional need to develop new ways of evaluating student writing, assessments mandated by postsecondary institutions and local, provincial, or national public education administrations directly challenge their personal and pedagogical integrity. What is the New Literacy practitioner to do when the time comes for his or her students to participate in a provincial writing assessment or write the Grade 12 English exam which must be passed by all students expecting high school diplomas (Willinsky & Bobie, 1986)? Such wide-scale assessments usually contain two components: a multiple-choice test, which might test reading comprehension or decontextualized writing skills

(knowledge of spelling, punctuation, word usage), and a writing sample, usually a piece of expository writing holistically marked. The testing context for the writing sample can, in fact, be designed so that students may use process writing approaches (Willinsky & Bobie, 1986, p. 5), although usually only the final product is holistically assessed. When wide-scale assessments are mandated, the friction between administrative demands for accountability and a teacher's need for professional autonomy is at its greatest. A possible solution lies in the teacher's ability to adapt to the immediate need (preparing students for the format of the test) without compromising long-term program goals.

The strategy is an old one: teach to the test, but with an awareness of a test's limitations and confidence that the writing abilities students have developed in the course of their instruction will serve them well during the exam. Nancie Atwell (1987), for example, states that

> I have no qualms about prepping kids for the state test, spending a few days prior to the exam talking about strategies for taking exams. I don't think prepping has that significant effect on kids' scores if the test calls for a writing sample. Kids can't learn how to write in preparation for a test, but they can learn how to better control the test situation. (p. 143)

Clearly, she is not intimidated by this type of assessment (as I believe teachers can be), and perhaps some of her own confidence is a powerful motivator to her students. Her statement also points to an unusual application of the New Literacy intention that

students control their own literacy; although a demonstration of literacy is being demanded by an outside agency, students are being taught how to bring this situation under their own control through knowledge of what the exam tests for.

Control can also be exerted in ways contrary to New Literacy intentions. Teachers can quite deliberately coach students to skew written responses in certain directions in order that their text will align with perceptions of how it might be evaluated. Willinsky and Bobie's 1986 study of the Grade 12 Alberta Diploma exams provides an interesting example of this type of control. In the academic version of these exams, students are asked to start with a personal response to two brief pieces of literature, arriving finally at a piece of expository prose on a literary theme. Allan Bobie, participant-observer in the study, draws upon his experiences as a marker of such exams in order to coach his students on the shaping of their responses. He teaches students to "keep their own first responses in check, in favor of what they would project an adult/teacher might welcome as a personal response from a responsible student" (Willinsky & Bobie, 1986, p. 5). Although coaching is a central metaphor for teacher practice of the New Literacy, the type of coaching undertaken in this case is quite unlike New Literacy theory and practice.

Nevertheless, I believe that teachers usually have good intentions for preparing students in this manner: they want their students to do well in these compulsory exercises of administrative authority, even if they might not agree with the intentions behind such tests. But the need to prepare students for exams can, and does, lead to teachers changing their instructional priorities in order that students have sufficient

practice in the discourse skills and conventions and testing situations which these tests call upon. Although knowledge of these skills is useful for successful test-writing, the danger is that classroom instruction can be so dominated by a focus on these areas that other equally valuable areas of literacy training suffer (Willinsky & Bobie, 1986, pp. 8-11). In the previous section of this paper, I stated that grading is the exercise of teacher power over the student. Wide-scale assessment programs can exercise a similar tyranny over teachers.

**The Politics of Assessment: A Final Reflection**

Although the New Literacy intends that teachers empower students with the ability to develop greater control, critical awareness, and independence as writers (Knoblauch & Brannon, 1984; Sommers, 1982; Atwell, 1987; Willinsky, 1990), I believe that it can enable teachers in a similar way. If teachers and students are more willing to share power and responsibility, both for learning and teaching, a shift in the direction of power takes place: power is no longer "top-down," but lateral, negotiated between the collaborators in literacy learning. If teachers are willing to accommodate a shift in the balance of classroom power, they must now challenge those who have wielded power over them in the form of administrative boardroom decisions to implement provincial, state, or national testing of student literacy abilities.

The movement toward such programs is slow but relentless. The diploma exams which Alberta's graduating students presently write are the final step in a provincial testing program which originally began as a random testing of one pupil in ten (Sarich,

1991, p. A1). *Answering the Challenge* (1990), a highly controversial policy document outlining the future development of high school education in Manitoba, plans for yearly curriculum assessment on the grounds that "common perceptions indicate that the school system is not graduating literate and knowledgeable students" (Manitoba Education and Training, p. 23).

Perhaps the most telling sign of things to come is Canada's School Achievement Indicators Project, a national testing program of the Council of Ministers of Education to be initiated in 1993. Curriculum differs from province to province, but preliminary examination of working documents for the test would indicate that the same standardized test will be offered throughout the country; for this reason, it can hardly serve as an instrument of curricular assessment. Furthermore, administering the same test throughout the country ignores the ethnic and demographic diversity of Canada's student population. Although present plans indicate that schools and students to be tested will be selected on a random basis (Sarick, 1991, p. A5), the experience of the Alberta and Manitoba provincial assessments suggests that random samplings can easily become mandated testing of all students at entire grade levels.

Perhaps the most insidious intention of this Canadian national testing project is that it will allow literacy performance comparisons to be made amongst provinces, schools, and school divisions. What is the intention of these comparisons? Will they reassure "employers who [want] to know that a high school diploma is the same across the country" (Sarick, 1991, p. A5)? Will they truly inform the public, which "wants to know if the schools are delivering" (Sarick, 1991, p. A1)? More importantly, what will be

done if great disparities emerge? Education is presently a provincial concern; will the results of these initiatives lead to a national Canadian curriculum? I hope not. The New Literacy offers great promise of pedagogical autonomy for the teacher and individual opportunity for students. Both will be undermined, if not destroyed, by a test-driven curriculum.

It can be argued that teachers may react pragmatically: even if they subscribe to the philosophy of the New Literacy, they can prepare students for the test and then return to the "real" curriculum. But if testing inevitably influences teaching (Willinsky & Bobie, 1986), then the present challenge of fitting test preparation into the classroom agenda will seem minor indeed. Teachers of the New Literacy have found the personal authority to challenge a variety of long-standing traditions in the teaching of reading and writing through their explorations of theory, their experiential knowledge of classroom realities, and reflective, active response to both. If teachers bring this same knowledge to bear on this most crucial issue of assessment, I think it can empower us to challenge equally long-standing traditions in the wide-scale evaluation of student ability and achievement.

I would be naive in believing that personal authority is a sufficient condition for an entire profession's empowerment. Power is also necessary, and as Miles Myers (1981) indicates in his essay on "The Politics of Minimum Competency," "there is a difference between power and authority. Power is what is achieved by constituencies--by counting heads and organizing large numbers" (p. 173). National organizations such as the Canadian Council of Teachers of English and the National Council of Teachers of

English can serve as a major constituency for English teachers in both countries if teachers are willing to join forces to make their concerns heard. By keeping their members aware of current trends and issues in literacy education, both CCTE and NCTE play an important role in ongoing professional development. Through their power to lobby at local and state or provincial levels, they can also exert the type of power which comes from an established constituency.

Myers is not naive, though. He points out that while many "decisions in education today are legitimized by counting heads and adding up constituencies," "some important decisions are legitimized by authority, not power, by appeals to expertise in a given area of study, by special knowledge through scholarship" (p. 173). Myers suggests that teachers can develop authority through observational research which becomes case study. Case studies are to be shared in order to exchange information, stimulate discussion and inquiry, and establish a published body of knowledge, much as the medical profession has done. Not surprisingly, Myers also states that authority comes from knowledge developed from models and theory, from the materials of classroom practice as well as from relevant research literature.

In making this point, I believe that Myers focuses on a crucial aspect of professional education for teachers. Even as they complete their final year of certification, novice teachers are often acutely aware of the limits of their knowledge; they undertake their school-based teaching assignments with rather superficial understanding of the theory behind the practice they undertake. Knowledge builds as a result of practical experience, but if it is not augmented by an awareness that theory

supports practice and that theory continues to change and develop, I believe that a teacher's potential for professional growth and extension will be limited. Faculties of education must cultivate in their students the habit of scholarship: of reading recent scholarship in one's field, of knowing which resources to which one can turn for reference or up-to-date information, in short, of continuing to learn.

For many years, I taught and graded as I had been taught and graded, largely because I was ignorant of the existence of alternative models As my doubts developed, I could neither disprove nor substantiate them because I had no idea as to where I might turn for information. Furthermore, even when doubts began to surface, the teaching situation in which I found myself led me to dismiss these doubts as personal idiosyncrasy rather than accept them as part of a genuine inner voice raising valid questions. In the presence of institutional authority, it is often hard to honor personal authority.

Fifteen years ago, evaluation of student writing was a very different issue for me. When I graded a paper for the first time, I was a student; now I am both a student and a teacher, although for the first time in my career, I have no papers to grade. Their temporary loss has been my gain. Only now have I been able to stand back and assess the interconnection between my theoretical understandings of the processes of reading and writing, my perceptions of authority in teachers, students, and texts, and my experiences as as reader, writer, teacher, and student of the New Literacy. Evaluation is more than picking up a red pen and assigning a letter or a number to a piece of writing. It is an issue which is moral, political, pedagogical, philosophical, and profoundly personal.

**References**

Applebee, Arthur N., Langer, Judith A., Jenkins, Lynn B., Mullis, Ina V. S., & Foertsch, Mary. (1990). *Learning to write in our nation's schools: Instruction and achievement in 1988 at grades 4, 8, and 12.* Princeton, NJ: National Assessment of Educational Progress.

Atwell, Nancie. (1987). *In the middle.* Portsmouth, NH: Boynton/Cook.

Baumlin, James S., & Baumlin, Tita French. (1989). Paper grading and the rhetorical stance. In B. Lawson, S. Ryan, & W. R. Winterowd (Eds.), *Encountering student texts* (pp. 171-82). Urbana, IL: National Council of Teachers of English.

Beach, Richard. (1989). Evaluating writing to learn: Responding to journals. In B. Lawson, S. Ryan, & W. R. Winterowd (Eds.), *Encountering student texts* (pp. 183-98). Urbana, IL: National Council of Teachers of English.

Beaven, Mary H. (1977). Individualized goal setting, self-evaluation, and peer evaluation. In C. Cooper & L. Odell (Eds.), *Evaluating writing: Describing, measuring, judging* (pp. 135-56). Urbana, IL: National Council of Teachers of English.

Bleich, David. (1975). *Readings and feelings: An introduction to subjective criticism.* Urbana, IL: National Council of Teachers of English.

Brannon, Lil, & Knoblauch, C. H. (1982). On students' rights to their own texts: A model of teacher response. *College Composition and Communication,* 33(2), 157-66.

Canadian Council of Teachers of English. (1985). *Evaluation policy.*

Charney, Davida. (1984). The validity of using holistic scoring to evaluate writing: A critical overview. *Research in the Teaching of English,* 18(1), 65-81.

Chater, Pauline. (1984). *Marking and assessment in English.* London: Methuen.

Cooper, Charles R. (1977). Holistic evaluation of writing. In C. Cooper & L. Odell (Eds.), *Evaluating writing: Describing, measuring, judging* (pp. 3-31). Urbana, IL: National Council of Teachers of English.

Cooper, Charles R. (Ed.). (1981). *The nature and measurement of competency in English.* Urbana, IL: National Council of Teachers of English.

Cooper, Charles R., & Odell, Lee. (1977). *Evaluating writing: Describing, measuring, judging.* Urbana, IL: National Council of Teachers of English.

Corder, Jim W. (1989). Asking for a text and trying to learn it. In B. Lawson, S. Ryan, & W. R. Winterowd (Eds.), *Encountering student texts* (pp. 89-97). Urbana, IL: National Council of Teachers of English.

Crowley, Sharon. (1989). On intention in student texts. In B. Lawson, S. Ryan, & W. R. Winterowd (Eds.), *Encountering student texts* (pp. 99-110). Urbana, IL: National Council of Teachers of English.

Diederich, Paul B. (1974). *Measuring growth in English.* Urbana, IL: National Council of Teachers of English.

Ede, Lisa. (1989). On writing reading and reading writing. In B. Lawson, S. Ryan, & W. R. Winterowd (Eds.), *Encountering student texts* (pp. 147-57). Urbana, IL: National Council of Teachers of English.

Flynn, Elizabeth. (1989). Learning to read student papers from a feminine perspective. In B. Lawson, S. Ryan, & W. R. Winterowd (Eds.), *Encountering student texts* (pp. 49-58). Urbana, IL: National Council of Teachers of English.

Freedman, Sarah Warshauer. (1979). Why do teachers give the grades they do? *College Composition and Communication,* 30(2), 161-64.

Freedman, Sarah Warshauer. (1981). Influences on evaluations of expository essays: Beyond the text. *Research in the Teaching of English,* 15(1), 245-55.

Freedman, Sarah Warshauer. (1987). *Response to student writing.* Urbana, IL: National Council of Teachers of English.

Freedman, Sarah Warshauer, & Robinson, William S. (1982). Testing proficiency in writing at San Francisco State University. *College Composition and Communication,* 33(4), 393-98.

Fulwiler, Toby (Ed.). (1987). *The journal book.* Portsmouth, NH: Boynton/Cook.

Gere, Anne Ruggles. (1980). Written composition: Toward a theory of evaluation. *College English,* 42(1), 48-58.

Gilbert, Mike. (1990). Responding to writers: What I value. *English Journal,* 79(5), 49-51.

Harris, Winifred Hall. (1977). Teacher response to student writing: A study of the response patterns of high school English teachers to determine the basis for teacher judgment of student writing. *Research in the Teaching of English,* 11(2), 175-85.

Hoetker, James. (1982). Essay examination topics and students' writing ability. *College Composition and Communication,* 33(4), 377-92.

Irmscher, William F. (1979). *Teaching expository writing.* New York: Holt Rinehart Winston.

Knoblauch, C. H., & Brannon, Lil. (1984). *Rhetorical traditions and the teaching of writing.* Portsmouth, NH: Boynton/Cook.

Lawson, Bruce, & Ryan, Susan Sterr. (1989). Interpretive issues in student writing. In B. Lawson, S. Ryan, & W. R. Winterowd (Eds.), *Encountering student texts* (pp. vii-xvii). Urbana, IL: National Council of Teachers of English.

Lloyd-Jones, Richard. (1977). Primary trait scoring. In C. Cooper & L. Odell (Eds.), *Evaluating writing: Describing, measuring, judging* (pp. 33-66). Urbana, IL: National Council of Teachers of English.

Manitoba Education. (1987). *English language arts 9-12.* Winnipeg, Manitoba.

Manitoba Education and Training. (1990). *Answering the challenge.* Winnipeg, Manitoba.

Manitoba Teachers' Society. (1989). *Policy handbook 1989 edition.* Winnipeg, Manitoba.

Marshall, James D. (1987). The effects of writing on students' understanding of literary texts. *Research in the Teaching of English,* 21(1), 30-63.

Murray, Patricia Y. (1989). Teachers as readers, readers as teachers. In B. Lawson, S. Ryan, & W. R. Winterowd (Eds.), *Encountering student texts* (pp. 73-85). Urbana, IL: National Council of Teachers of English.

Myers, Miles. (1981). The politics of minimum competency. In C. Cooper (Ed.), *The nature and measurement of competency in English* (pp. 165-74). Urbana, IL: National Council of Teachers of English.

National Council of Teachers of English. (1990). On the development and dissemination of alternative forms of assessment. Resolution 1, *Minutes of the Annual Business Meeting.*

Nelson, Jennie. (1990). This was an easy assignment: Examining how students interpret academic writing tasks. *Research in the Teaching of English,* 24(4), 362-96.

Nold, Ellen W., & Freedman, Sarah W. (1977). An analysis of readers' responses to essays. *Research in the Teaching of English,* 11(2), 164-74.

Probst, Robert E. (1988). *Response and analysis: Teaching literature in junior and senior high school.* Portsmouth, NH: Boynton/Cook.

Sarick, Lila. (1991, May 16). To test or not to test: That is the question. *The Globe and Mail,* pp. A1; A5.

Shaughnessy, Mina P. (1977). *Errors and expectations.* New York: Oxford University Press.

Sommers, Nancy. (1982). Responding to student writing. *College Composition and Communication,* 33(2), 148-56.

Tierney, Robert J., & Pearson, P. David. (1983). Toward a composing model of reading. In J. Jensen (Ed.), *Composing and comprehending* (pp. 33-45). Urbana, IL: National Council of Teachers of English.

Tyler, Ralph W. (1949). *Basic principles of curriculum and instruction.* Chicago: University of Chicago Press.

University of Winnipeg. (1991). *Faculty of arts and science 1991-92 calendar.* Winnipeg, Manitoba.

*Webster's New Collegiate Dictionary.* (1981). Toronto: Thomas Allen and Son.

White, Edward M. (1984). Post-structural literary criticism and the response to student writing. *College Composition and Communication,* 35(2), 186-95.

Willinsky, John. (1990). *The new literacy.* London: Routledge.

Willinsky, John, & Bobie, Allan. (1986). *When tests dare to be progressive.* ED 278 964.

Winterowd, W. Ross. (1989). The drama of text. In B. Lawson, S. Ryan, & W. R. Winterowd (Eds.), *Encountering student texts* (pp. 21-33). Urbana, IL: National Council of Teachers of English.

## NCTE Concept Paper Series

NCTE's series of *Concept Papers* was established to present innovative thinking about the field of English language arts education in a relatively informal, quickly produced format. Designed for works that are too long for publication in professional journals, but still much shorter than standard monographs, *Concept Papers* explore theoretical and practical issues in English literacy education at all teaching levels.

***NCTE Concept Papers available***

1. *Guidelines for Judging and Selecting Language Arts Textbooks: A Modest Proposal*; Timothy Shanahan, Lester Knight, et al.
2. *Doublespeak: A brief history, definition, and bibliography, with a list of award winners, 1974-1990*; Walker Gibson and William Lutz
3. *Alternatives in Understanding and Educating Attention-Deficit Students: A Systems-Based Whole Language Perspective*; Constance Weaver
4. *A Goodly Fellowship of Writers and Readers*; Richard Lloyd-Jones



BEST COPY AVAILABLE